Action Recognition by Graph Embedding and Temporal Classifiers

Ehsan Zare Borzeshi

Faculty of Engineering and Information Technology

University of Technology, Sydney

A dissertation submitted for the degree of

Doctor of Philosophy

May 2014

Certificate of Authorship and Originality

Title: **Action Recognition by Graph Embedding and Temporal Classifiers**
Author: **Ehsan Zare Borzeshi**
Date: **May 1, 2014**
Degree: **PhD**

I certify that the work in this dissertation has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the dissertation has been written by me. Any help that I have received in my research work and the preparation of the dissertation itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the dissertation.

Signiture of author

I would like to dedicate this dissertation to my beloved wife

Shima

for her unending love and support.

Acknowledgements

Many people have contributed in various ways to make this PhD study an exciting and memorable journey. Just to name a few:

I would like to acknowledge the lively and stimulating atmosphere in our group. The lunches, discussions at the coffee shops and evenings out: I have enjoyed all of them. Several persons deserve a special mention. I gratefully acknowledge all of my current and former group members: Dr. Oscar Perez Concha, Ava Bargi, Shaukat Abidi, Jaime Andres Garcia, Dr. Richard Yi Da Xu; my friends: Prof Federico Girosi and Dr. Majid Nazem; and members of the iNext research centre at UTS. Furthermore, I owe sincere gratitude to all members of the Center for Research in Computer Vision at UCF: Afshin Dehghan, Shayan Modiri and Amir Roshan Zamir; and Dr. Kasper Riesen at University of Applied Sciences and Arts Northwestern Switzerland.

I would like to express my utmost gratitude to my external advisors, Professor Mubarak Shah and Professor Horst Bunke for all their invaluable help and advice during my studies.

I would like to acknowledge ARC and my colleagues in the "Airports of the Future" project. This work has been supported by the Australian Research Council (ARC) under Linkage Projects Scheme "LP0990135".

I would like to thank my dissertation reviewers for finding the time to read and comment on this work.

I would like to thank my parents, Mohammad and Marzieh, for giving birth to me at the first place, teaching me that it is no disgrace to work hard (which has proven to be a useful lesson) and supporting me spiritually throughout my life. I also thank them for their faith in me and allowing

me to be as ambitious as I wanted. It was under their watchful eye that I have gained so much drive and ability to tackle challenges head on. Furthermore, I thank my sister and brother, Elaheh and Amirhossein, for their love and support. I am so lucky to have all of you as my family.

I would also like to thank my wife's parents, Hossein and Tahereh, for first letting me take her hand in marriage, and for their extreme kindness and invaluable support to date. They are my best friends and beloved family and their interest and support has been tremendously valuable and appreciated indeed.

A huge thanks to my supervisor Professor Massimo Piccardi for supporting me during my PhD, for being patient and for being critical about my work. I admire his brilliance, determination, and hunger for knowledge. I have very much appreciated his pleasant way of supervising, even though I have never mentioned this explicitly. He has been an invaluable source of insights that any PhD student would love to know but which have never been written down. Massimo has supported me academically and emotionally through the rough road to finish this thesis. I feel deeply indebted to Massimo and a huge portion of the knowledge and confidence that I have today is due to him.

This list would be incomplete without expressing the most wholehearted gratitude to my life angel and love, Shima. Her support has been unconditional all these years; she has given up many things for me to finish my study; she has cherished with me every great moment and supported me whenever I needed it. Her positive spirit, unwavering love, patient endurance and tolerance of my occasional vulgar mood throughout the last ten years deserve so much more than just a "thank you". Not only now, but also in the years to come.

Abstract

With the improved accessibility to an exploding amount of video data and growing demand in a wide range of video analysis applications, video-based action recognition becomes an increasingly important task in computer vision. Unlike most approaches in the literature which rely on bag-of-feature methods that typically ignore the structural information in the data, in this monograph we incorporate the spatial relationship and the time stamps in the data in the recognition and classification processes.

We capture the spatial relationships in the subject performing the action by representing the actor's shape in each frame with a graph. This graph is then transformed into a vector of real numbers by means of prototype-based graph embedding. Finally, the temporal structure between these vectors is captured by means of sequential classifiers. The experimental results on a well-known action dataset (KTH) show that, although the proposed method does not achieve accuracy comparable to that of the best existing approaches, these embedded graphs are capable of describing the deformable human shape and its evolution over time.

We later propose an extended hidden Markov model, called the hidden Markov model for multiple, irregular observations (HMM-MIO), capable of fusing spatial information provided by graph embedding and the textural information of STIP descriptors. Experimental results show that recognition accuracy can be significantly improved by combining the spatio-temporal features with the structural information obtaining higher accuracy than from either separately. Furthermore, HMM-MIO is applied to the task of joint action segmentation and classification over a concatenated version of the KTH action dataset and the challenging CMU multi-modal activity dataset. The achieved accuracies proved comparable to or higher

than state-of-the-art approaches and show the usefulness of the proposed model also for this task.

The next and most remarkable contribution of this dissertation is the creation of a novel framework for selecting a set of prototypes from a labelled graph set taking class discrimination into account. Experimental results show that such a discriminative prototype selection framework can achieve superior results, not only for the task of human action recognition, but also in the classification of various structured data such as letters, digits, drawings, fingerprints compared to other well-established prototype selection approaches.

Lastly, we change our focus from the forementioned problems to the recognition of complex event, which is a recent area of computer vision expanding the traditional boundaries of visual recognition. For this task, we have employed the notion of concept as an alternative intermediate representation with the aim of improving event recognition. We model an event by a hidden conditional random field and we learn its parameters by a latent structural SVM approach. Experimental results over video clips from the challenging TRECVID MED 2011 and MED 2012 datasets show that the proposed approach achieves a significant improvement in average precision at a parity of features and concepts.

Contents

List of Figures